



AVANT-PROPOS

Cette Troisième Rencontre fait suite aux Journées intitulée « La fouille dans les données par la Méthode Statistique Implicative » qui se sont déroulées à l'Institut de Formation des Maîtres de Caen les 23-24 Juin 2000. Elles étaient organisées par Marc Bailleul et Régis Gras. Ces Journées ont été suivies par le Colloque « O metodo estatístico implicativo utilizado em estudos qualitativos de Régis Gras de associação. Contribuição à pesquisa em Educação » à l'Université PUC de Sao Paulo 9-11 Juillet 2003, organisé par Saddo Ag Almouloud. Ainsi, suite à l'intérêt porté à ces Journées, c'est la 3^{ème} fois que se rassemblent des chercheurs représentant de nombreuses nationalités sur le thème de l'Analyse Statistique Implicative (A.S.I.).

Rappelons que la "fouille dans des données" (encore appelée "Knowledge Discovery in Databases" ou encore "Data Mining" dans la littérature anglo-saxonne) part, en général, du croisement de sujets (ou objets) et de variables (propriétés ou attributs) binaires, ordinales ou numériques. Son objectif majeur consiste à conjecturer des modèles basés sur des relations quantitatives ou qualitatives et des structures induites à partir des données. Différentes méthodes, comme l'Analyse Factorielle des Correspondances (A.F.C.), la Classification Ascendante Hiérarchique (C.A.H.), sont communément utilisées pour de telles fouilles dans des données. Parmi elles, l'Analyse Statistique Implicative (A.S.I.) vise l'extraction de connaissances, d'invariants, de règles inductives non symétriques consistantes, et accorde une mesure à des propositions du type « quand a est choisi, on a tendance à choisir b ». Pour cela, elle quantifie la qualité de ces règles sur la base statistique d'un nombre significatif de contre-exemples où la règle n'est pas vérifiée et où certains déséquilibres cardinaux sont observés parmi les exemples et les contre-exemples à l'implication et à sa contraposée. Le logiciel, dénommé CHIC (Classification Hiérarchique Implicative et Cohésitive), développé au début des années 90 par Saddo Ag Almouloud et depuis par Raphaël Couturier, permet :

- de traiter différents types de variables (binaires, modales, fréquentielles, intervalles, floues),
- de quantifier la significativité des valeurs attribuées à la qualité, la consistance de la règle associée, de classes ordonnées de règles, à la typicalité et la contribution des sujets ou de catégories de sujets à certaines règles,
- de représenter, par un graphe, pour un seuil de qualité choisi, des chemins de règles et, par une hiérarchie, des règles sur des règles que l'on appelle aussi règles généralisées,
- de supprimer, d'ajouter, de conjoindre des variables.
- de représenter par une hiérarchie ascendante les classes de similarité des variables.

Pendant ces trois Journées, des communications de chercheurs de plusieurs pays (cette année : Brésil, Chypre, Espagne, France, Grèce, Italie, Suisse et Viet-Nam) vont



rendre compte de travaux et résultats significatifs obtenus dans différents domaines par une "fouille" dans leurs propres données. Ces résultats, précédés de la conférence de Régis Gras sur un rappel des concepts fondamentaux de l'A.S.I., font l'objet des communications de la 1^{ère} Journée. L'A.S.I. a permis de faire apparaître, de façon complémentaire et comparative à des méthodes factorielles ou taxonomiques, des règles en bio-informatique, en psychologie sociale, en éducation, etc.. La conférence de Filippo Spagnolo en ouverture de la 2^{ème} Journée, met en parallèle les méthodologies respectives en A.F.C. et en A.S.I.. Les communications qui suivent cette conférence illustrent souvent cette comparaison à partir de situations observées en Didactique des Mathématiques. La 3^{ème} Journée, ouverte par la conférence de Pascale Kuntz sur la construction d'un modèle original de hiérarchie ascendante, non symétrique en A.S.I., présente quelques concepts théoriques statistiques de règles d'association en relation plus ou moins étroite avec l'A.S.I..

Au cours de ces Rencontres, l'intérêt d'une alternance équilibrée de communications théoriques ou appliquées et de travaux pratiques sur CHIC, est toujours souligné favorablement à la fois par les chercheurs chevronnés et par les jeunes chercheurs. Car ce sont les démarches dévoilées de conceptualisation, d'analyse et d'interprétation qui sont les plus utiles aux uns et aux autres. Les débats qui accompagnent chacune des interventions sont l'occasion de préciser et de spécifier certains points permettant d'éclairer et de guider des recherches en cours ou à venir.

Nous avons tenu à regrouper dans le même lieu, l'Hôtel Villagio Città del Mare à Terrasini, près de la mer, de Palermo, et de son aéroport, à la fois les activités scientifiques, l'hébergement et la restauration. Les conditions de travail, le cadre et le confort de cet Hôtel ne peuvent que conduire à la convivialité et à la continuité des échanges.

Nous tenons à remercier les différents organismes qui nous ont apporté un soutien à la tenue de ces Journées sur les plans scientifique, financier et logistique :

l'Association pour la Recherche en Didactique des Mathématiques (ARDM),

l'Association Extraction et Gestion des Connaissances (EGC),

l'Université de Palermo, en la personne de son Recteur, le Professeur Giuseppe Silvestri,

la Faculté des Sciences de la Formation, en la personne du Professeur Patrizia Lendinara,

l'Ecole Polytechnique de l'Université de Nantes,

le Laboratoire d'Informatique de Nantes (LINA),

le Gruppo di Ricerca sul l'Insegnamento della Matematiche (GRIM)

la Société PerformanSe SA, Carquefou- Nantes,

ou leur parrainage scientifique :

La Società Italiana di Statistica (SIS)

la Société française de Statistique (SFdS)

la Société Francophone de Classification (SFC),

l'International Association for Statistical Education (IASE)

Le Comité d'organisation



PREFAZIONE

Questo terzo incontro fa seguito alle giornate su « L'indagine dei dati con il Metodo di Statistica Implicativa » che si sono svolte presso l'Istituto di Formazione degli Insegnanti di Caen, 23-24 Giugno 2000. Organizzate da Marc Bailleul e Régis Gras. Queste giornate sono state seguite dal Convegno « O metodo estatístico implicativo utilizado em estudos qualitativos de Régis Gras de associação. Contribuição à pesquisa em Educação » presso l'Università PUC di San Paolo, 9-11 Luglio 2003, organizzato da Saddo Ag Almouloud. Facendo seguito all'interesse avuto in queste convegni, ci si incontra sul tema dell'A.S.I. per la terza volta riunendo ricercatori rappresentanti di numerose nazioni.

Ricordiamo che una « indagine » dei dati (chiamata anche "Knowledge Discovery in Databases" o "Data Mining" nella letteratura anglosassone) parte, in generale, dall'incrocio di soggetti (o oggetti) e di variabili (proprietà o attributi) binari, ordinali o numerici. Il suo obiettivo principale consiste nel congetturare dei modelli basati su relazioni quantitative o qualitative e di strutture indotte a partire da dati. Differenti metodi, come l'Analisi Fattoriale delle Corrispondenze (A.F.C.), la Classificazione Ascendente Gerarchica (C.A.H.), sono utilizzate comunemente per queste indagini di dati. Tra queste, l'Analisi Statistica Implicativa (A.S.I.) si rivolge all'estrazione delle conoscenze, di invarianti, di regole induttive non simmetriche consistenti, e fornisce una misura a delle proposizioni del tipo “quando **a** è scelta, si ha la tendenza a scegliere **b**”. Per questo, quantifica la qualità di queste regole sulla base statistica di un numero significativo di contro esempi dove la regola non è verificata e dove alcuni disequilibri cardinali sono osservati attraverso gli esempi ed i contro esempi all'implicazione ed alla sua contrapposta. Il software CHIC (Classificazione Gerarchica Implicativa e Coesiva), sviluppato nei primi anni '90 da Saddo Ag Almouloud ed in seguito da Raphaël Couturier, permette:

- di trattare differenti tipi di variabili (binarie, modali, frequenziali, intervalli fuzzy);
- di quantificare la significatività dei valori attribuiti alla qualità, la consistenza della regola associata, di classi ordinate di regole, alla tipicità e la contribuzione di soggetti o di categorie di soggetti a certe regole;
- di rappresentare, con un grafo, per una determinata soglia di qualità scelta, dei cammini di regole e, da una gerarchia, delle regole su delle regole che sono chiamate regole generalizzate;
- di sopprimere, di aggiungere, di congiungere delle variabili;
- di rappresentare con una gerarchia ascendente le classi di similarità di variabili.

Nel corso di queste tre giornate, vi saranno comunicazioni di ricercatori di diversi paesi (quest'anno: Brasile, Cipro, Spagna, Francia, Grecia, Italia, Svizzera e Vietnam) si confronteranno su lavori e risultati significativi ottenuti nei differenti campi con una



“indagine” sui propri dati. Questi risultati, preceduti dalla conferenza di Régis Gras su di un richiamo ai concetti fondamentali dell’A.S.I., sono l’oggetto delle comunicazioni della prima giornata. L’A.S.I. ha permesso di far vedere, in maniera complementare e comparativa con dei metodi fattoriali o tassonomici, delle regole bio-informatiche, in psicologia sociale, in educazione, etc... La conferenza di Filippo Spagnolo ad apertura della seconda giornata mette in parallelo le metodologie dell’A.F.C. e dell’A.S.I. Le comunicazioni che seguiranno questa conferenza illustrano spesso questa comparazione a partire da situazioni osservate in Didattica della Matematica. La terza giornata, aperta dalla conferenza di Pascale Kuntz sulla costruzione di un modello originale di gerarchia ascendente, non simmetrica nell’A.S.I., presenta alcuni concetti teorici statistici di regole di associazione in relazione più o meno stretta con l’A.S.I.

Nel corso di questi incontri, l’interesse di una alternanza equilibrata di comunicazioni teoriche o applicate e di lavori paratici con lo CHIC, é sempre sottolineato favorevolmente sia dai ricercatori più anziani che quelli giovani. Perché queste sono i passi da compiere per la concettualizzazione, per l’analisi e per l’interpretazione che sono molto utili sia agli uni che agli altri.. I dibattiti che accompagnano ciascuno degli interventi sono occasione per precisare e per specificare certi punti che permettono di chiarire e di guidare i ricercatori di oggi e di domani.

Abbiamo voluto raggruppare nello stesso luogo, Hôtel Villaggio Città del Mare a Terrasini, vicino al mare, vicino a Palermo ed al suo aeroporto, al tempo stesso, attività scientifiche, alloggio e ristorazione. Le condizioni di lavoro, il quadro e il conforto di questo Hotel non possono che condurre ad una convivialità e ad una continuità di scambi.

Ringraziamo i diversi organismi che ci hanno portato un sostegno alla riuscita di queste giornate sul piano scientifico, finanziario e logistico:

l’Association pour la Recherche en Didactique des Mathématiques (ARDM),

l’Association Extraction et Gestion des Connaissances (EGC),

l’Università di Palermo, nella persona del Rettore Prof. Giuseppe Silvestri,

la Facoltà di Scienze della Formazione, nella persona della Preside Prof.ssa Patrizia Lendinara,

l’Ecole Polytechnique de l’Université de Nantes,

le Laboratoire d’Informatique de Nantes (LINA),

le Gruppo di Ricerca sull’Insegnamento della Matematiche (GRIM)

la Société PerformanSe SA, Carquefou- Nantes,

e gli sponsors scientifici :

La Società Italiana di Statistica (SIS)

la Société française de Statistique (SFdS)

la Société Francophone de Classification (SFC),

l’International Association for Statistical Education (IASE)

Il Comitato di Organizzazione



***Comité Scientifique et de Programme,
 3ème Rencontre A.S.I.
 Palerme, 6-8 Octobre 2005***

Saddo AG ALMOULOUD, Université P.U.C. de Sao Paulo,
Marc BAILLEUL, I.U.F.M. de Caen,
Serge BAQUEDANO, Société PerformanSe, Nantes-Carquefou,
Henri BRIAND, LINA, Université de Nantes,
Raphaël COUTURIER, LIFC - IUT Belfort,
Edwin DIDAY, CEREMADE, Paris Dauphine,
Athanasios GAGATSIS, Université de Chypre,
Ornella GIAMBALVO, Università di Palermo,
Robin GRAS, Swiss Institute of Bioinformatics, Genève,
Fabrice GUILLET, LINA, Université de Nantes
Yves KODRATOFF, LRI-C.N.R.S., Paris,
Pascale KUNTZ, LINA, Université de Nantes,
Ludovic LEBART, ENST-CNRS, Paris,
Israël-César LERMAN, IRISA, Université de Rennes 1,
Anna Maria MILITO, Università di Palermo,
Guy NOEL, Université de Mons,
Pilar ORUS, Universitat Jaume-I, Castellon
Maria Gabriella OTTAVIANI, Università di Roma,
Maria A. PANNONE, Università di Perugia,
Anna PARROCO, Università di Palermo,
Maria POLO, Università di Cagliari,
Jean-Claude REGNIER, EA Savoirs, Diversité et Professionnalisation, Université Lyon 2,
Gilbert RITSCHARD, Département Econométrie, Université de Genève,
Filippo SPAGNOLO, GRIM, Università di Palermo,
Djamel ZIGHED, ERIC, Université Lyon 2

Régis GRAS, LINA, Université de Nantes, **Président** du Comité Scientifique et de Programme.

Comité d'organisation

Jérôme DAVID, Université de Nantes
Benedetto Di PAOLA, Università di Palermo
Régis GRAS, Université de Nantes
Pascale KUNTZ, Université de Nantes

Filippo SPAGNOLO, Università di Palermo, **Président** du Comité d'organisation